Industrial Organization: Problem Set 1

Tanya Rajan and George Vojta

February 3, 2021

Note: We submit code in 2 separate files: **pset1_8and14** contains the code for Questions 8 and 14 while **pset1_9thru13** contains the code for Questions 9 through 13.

Consider a unit mass of consumers choosing over products $j \in \{0, 1, 2, \ldots, J\}$ in markets $t = 1, 2, \ldots, T$. Each consumer i 's indirect utility of purchasing product j is:

$$u_{ijt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \varepsilon_{ijt}$$

where x_{jt} is observable product characteristics and p_{jt} is price. ξ_{jt} is unobserved (by researchers) product characteristics. ε_{ijt} is unobserved (by researchers) product/consumer match value that is distributed iid as T1EV. Product 0 is the outside option. We normalize the utility of choosing the outside option as $u_{i0t} = \varepsilon_{i0t}$. Let θ denote (α, β)

1. Derive the aggregate market share s_{jt} for product j in market t

Here we are trying to derive the aggregate market share s_{jt} under a constant coefficient model. First we will prove two lemmas that will be useful later:

Lemma 0.1. *The maximum of a T1EV distribution is itself a T1EV distribution.*

Proof. Take some constant $c \in \mathbb{R}$ and suppose $u_{ij} \sim T1EV(\alpha_j)$. Then let us consider the value of $\mathbb{P}(c > \max_{j \in \mathcal{J}} u_{ij})$ where $\mathcal{J} = \{0, 1, \ldots, J\}$. We can rewrite this as:

$$\begin{aligned}
\mathbb{P}(c > \max_{j \in \mathcal{J}} u_{ij}) &= \Pi_{j=1}^{J} \mathbb{P}(c > u_{ij}) \\
&= \Pi_{j=1}^{J} \exp\{-\exp(-(c - \alpha_j))\} && \text{(from T1EV CDF)} \\
&= \exp\{-\sum_{j=1}^{J} \exp(-(c - \alpha_j))\} \\
&= \exp\{-\exp(-c) \sum_{j=1}^{J} \exp(\alpha_j)\} \\
&= \exp\left\{-\exp\left(-(c - \ln \sum_{j=1}^{J} \exp(\alpha_j))\right)\right\}
\end{aligned}$$

Notice that the last expression is CDF for a random variable distributed $T1EV\left(\ln \sum_{j=1}^{J} \exp(\alpha_j)\right)$. (In the textbook this is usually denoted as $\alpha = \ln \sum_{j=1}^{J} \exp(\alpha_j)$).

□

Lemma 0.2. *If $u_{ij} \sim T1EV(\delta_j)$ and $u_{ik} \sim T1EV(\delta_k)$, then $u_{ij} - u_{ik} \sim Logistic(\delta_j - \delta_k)$.*

We leave out the proof of this lemma, but it can be found in the textbook among other places.

Now let us turn back to the problem at hand. First let us define $\delta_j := x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$. So we can write $u_{ijt} = \delta_j + \varepsilon_{ijt}$. We know that ε_{ijt} are iid and distributed as T1EV. Further, we have infinitely many consumers who are identical in their preferences. So we know that the share of good j in the market is simply defined by the probability that consumer i chooses good j.

$$\begin{aligned}
\mathbb{P}(i \text{ choose } j) &= \mathbb{P}\left(u_{\text{ijt}} > u_{\text{ikt}}\right) \quad \forall k \in \mathcal{J}/\{j\} \\
&= \mathbb{P}(\underbrace{\delta_j + \varepsilon_{ijt}}_{\sim T1EV(\delta_j)} > \underbrace{\max_{k \neq j} \delta_k + \varepsilon_{ikt}}_{\sim T1EV(\ln \sum_{i \neq j} \exp(\delta_k))})
\end{aligned}$$

Note that we know the right hand side of the inequality is distributed T1EV by the first lemma we proved. Next we can subtract off the LHS of the inequality and use Lemma 2 to say:

$$\begin{aligned}
\mathbb{P}(i \text{ choose } j) &= \mathbb{P}(0 > \underbrace{\max_{k \neq j} \delta_k + \varepsilon_{ikt} - \delta_j - \varepsilon_{ijt}}_{\sim Logistic(\ln \sum_{k \neq j} \exp(\delta_k) - \delta_j)}) \\
&= \frac{1}{1 + \exp\{-(\ln \sum_{k \neq j} \exp(\delta_k) - \delta_j)\}}
\end{aligned}$$

Here we can note that $\exp\{\ln \sum_{k \neq j} \exp(\delta_k) - \delta_j\} = \frac{\sum_{k \neq j} \exp(\delta_k)}{\exp(\delta_j)}$. Then we can rewrite the logisitic CDF from above as:

$$\begin{aligned}
\mathbb{P}(i \text{ choose } j) &= \frac{1}{1 + \frac{\exp(\delta_j)}{\sum_{k \neq j} \exp(\delta_k)}} \\
&= \frac{\exp(\delta_j)}{\exp(\delta_j) + \sum_{k \neq j} \exp(\delta_k)} \\
&= \frac{\exp(\delta_j)}{\sum_{k \in \mathcal{J}} \exp(\delta_k)} \\
&= \frac{\exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})}{\sum_{k \in \mathcal{J}} \exp(x_{kt}\beta - \alpha p_{kt} + \xi_{kt})}
\end{aligned}$$

Moving forward we will refer to this probability as π_{jt}:

$$[\text{Share j :}] : \quad \pi_{jt} := \frac{\exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})}{\sum_{k \in \mathcal{J}} \exp(x_{kt}\beta - \alpha p_{kt} + \xi_{kt})}$$

Also note that across a set of individuals, probabilities become shares thus is sufficient to just solve for the probability above.

2. Derive the elasticity matrix for a market with J goods. Describe the IIA property and the associated concerns. What are some possible remedies?

The elasticities can be deriving using the expression $\frac{\partial \pi_{jt}}{\partial p_{ht}}\frac{p_{ht}}{\pi_{jt}}$. This is due to the fact that consumers can only purchase one good in the model, thus demand is the share. The own-price elasticities are when $j = h$. So we can take derivatives. For the own-price elasticity:

$$\begin{aligned}
\frac{\partial \pi_{jt}}{\partial p_{jt}} &= \frac{-\alpha \sum_k \exp(\delta_n)\exp(\delta_j) + \alpha \exp(\delta_j)\exp(\delta_j)}{\left(\sum_k \exp(\delta_k)\right)^2} \\
\frac{\partial \pi_{jt}}{\partial p_{jt}}\frac{p_{jt}}{\pi_{jt}} &= p_{jt}\frac{-\alpha \sum_k \exp(\delta_k)\exp(\delta_j) + \alpha \exp(\delta_j)\exp(\delta_j)}{\left(\sum_k \exp(\delta_k)\right)^2}\frac{\sum_{k\in\mathcal{J}}\exp(\delta_k)}{\exp(\delta_j)} \\
&= p_{jt}\frac{-\alpha \sum_k \exp(\delta_k) + \alpha \exp(\delta_j)}{\sum_k \exp(\delta_k)} \\
&= -\alpha p_{jt}(1-\pi_{jt})
\end{aligned}$$

For the cross-price elasticity of good j with respect to good h:

$$\begin{aligned}
\frac{\partial \pi_{jt}}{\partial p_{ht}} &= \frac{\alpha \exp(\delta_j)\exp(\delta_h)}{\left(\sum_k \exp(\delta_k)\right)^2} \\
\frac{\partial \pi_{jt}}{\partial p_{ht}}\frac{p_{ht}}{\pi_{jt}} &= p_{ht}\frac{\alpha \exp(\delta_j)\exp(\delta_h)}{\left(\sum_k \exp(\delta_k)\right)^2}\frac{\sum_{k\in\mathcal{J}}\exp(\delta_k)}{\exp(\delta_j)} \\
&= p_{ht}\frac{\alpha \exp(\delta_h)}{\sum_k \exp(\delta_k)} \\
&= \alpha p_{ht}\pi_{ht}
\end{aligned}$$

So the elasticity matrix is:

$$\begin{cases} -\alpha p_{jt}(1-\pi_{jt}) & \text{if } j = h \\ \alpha p_{ht}\pi_{ht} & \text{otherwise} \end{cases}$$

The IIA property in the logit model comes from the fact that if we take the ratio of the choice probabilites for two goods, we will get $\frac{\pi_{it}}{\pi_{ht}} = \frac{\exp\{\delta_j\}}{\exp\{d_h\}}$, which does not depend on the utility gained from any of the other alternatives other than j and h. However, in reality, the substitution patterns are not this simple. For example, if the price of a kids breakfast cereal (e.g. Cocoa Puffs) increases, consumers are more likely to substitute towards another kids cereal (e.g. Lucky Charms) than to substitute towards an adult cereal (e.g. Grape Nuts). However, in the basic logit model, the substitution patterns are determined by simply market prices and shares as shown in the elasticity equations above. If the three cereals in the examples above all have the same market share and the price of Cocoa Puffs increases, the logit model predicts that people substitute equally towards Lucky Charms and Grape Nuts.

Solutions to this problem include nesting products (e.g. kids vs. adult cereals) or using a fully random coefficient model that allows β to be a random variable. Alternatively, using multinomial probit with a Gaussian distribution assumption on the error terms will also allow correlations across alternatives.

> **3.** Suppose the researchers observe market shares s_{jt} in the data. Are observed zero market shares consistent with the model we have just described? If not, why and discuss some potential solutions.

No—zero market shares do not make sense in the logit model. If a product had zero market share, then the own-price elasticity of substitution would be equal to $-\alpha p_{jt}$ while the cross-price elasticity of substitution would be 0. Further, from the inversion equation in the next question, we see that the model would break when we try to estimate log(0).

$$
\begin{aligned}
\text{[Share j :]}: \quad \pi_{jt} &:= \frac{\exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})}{\sum_{k\in\mathcal{J}} \exp(x_{kt}\beta - \alpha p_{kt} + \xi_{kt})} \\
\Rightarrow \pi_{jt} = 0 &\iff \exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt}) = -\infty
\end{aligned}
$$

Which is clearly problematic. Solutions to this problem include assuming error in the demand through shocks s.t. true shares aren't infinity defining demand shocks as the following:

$$\phi = \eta(\mu_j) + \xi$$

As a function of the propensity score of not being selected (μ), ala Dube et al (2020). Solving out the model from there and by taking logs of shares it is possible to create a new moment for estimation (as in page 54 of the textbook, eq 3.2.21). If you do not want to incorporate random utility shocks the same paper has an example to abstract away from those by leveraging CES functional forms.

> **4.** Define mean utilities $\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$. Show that δ_{jt} can be written as a function of market shares by inversion. Discuss to what extent does the inversion depend on the functional form assumptions.

Using the given assumption (/ normalization) that $u_{i0t} = \varepsilon_{i0t}$, we can write that

$$\pi_{i0t} = \frac{1}{\sum_k \exp(\delta_k)}$$

Then if we take the ratio of the choice probabilities for good j to the outside option good and take logs, we get:

$$
\begin{aligned}
\frac{\pi_{ijt}}{\pi_{i0t}} &= \frac{\exp(\delta_j)}{\sum_k \exp(\delta_k)} \sum_k \exp(\delta_k) = exp(\delta_j) \\
\ln\left(\frac{s_{jt}}{s_{0t}}\right) &= \delta_j = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}
\end{aligned}
$$

We use the IIA structure to eliminate dependence on all other goods, the homogeneous consumer assumption so that market shares are equal to individual choice probabilities, and our normalization of the outside option to simplify further. In the case of random coefficients, the share equation is still invertible, but will require more computation (e.g. searching for the value of δ_j that minimizes the distance between predicted and observed market shares. So in some sense, the functional form assumptions here are buying us ease of inversion, but barring computational constraints, it is not necessary for inversion.

> **5.** Are the OLS estimates based on the computed mean utilities above consistent for demand parameters α and β? If not, propose solutions on how to get consistent estimates.

OLS estimates based on the utilities above are alone not consistent. We need to account for endogeneity of ξ_{jt}. So we have to propose an appropriate instrument. Cost variables such as input prices (that are excluded from x_{jt}), product characteristics of competitors (BLP instruments), or prices of the same good in other cities (Haussman instruments) may all provide valid sources of exogenous variation.

6. What is ξ in this setting? Discuss the validity of both Haussman and BLP instruments in this setting. Regardless of your answers, do not include these as additional instruments for computation.

The ξ term captures the mean valuation of the product characteristics that are unobserved to the econometrician but observed to the consumer. At the same time, ξ can also be interpreted as the supply side valuation of product characteristics. Clearly this term will be correlated with p_{jt} through the pricing decision of firms, which is why we need instruments.

Haussman instruments are the prices of the same good in a different city. The assumption needed to use these instruments are that while prices are correlated across cities due to the common marginal cost components, price in city B will not affect demand shifts in city A. BLP instruments make use of the valuation of the same product characteristics produced by other firms. The basic argument is that if you can isolate product characteristics that occur before the ξ term is realized then (they argue) that the price of that good depends on the product characteristics of their competitors as well but that those opponent characteristics are orthogonal to ξ. Written mathematically these assumptions buy you that:

$$\begin{aligned} x_{-j} &\perp \xi_j \\ x_{-j} &\not\perp p_j \end{aligned}$$

Which come directly from lecture.

7. Consider a firm that potentially sells multiple products in a market where price is set simultaneously by each firm. Derive the FONCs for the firm. Combine these FONC's for all firms for a fixed-point equation describing the equilibrium.

' Consider the profit for a firm with $j \in J$ goods on the market. Profit can be decomposed into mark-up times demand:

$$\begin{aligned} \Pi_{ft} &= \sum_j (p_{jt} - mc_{jt}) \pi_{jt}(p, x, \xi, \theta) \\ &= \sum_j N \left((p_{jt} - mc_{jt}) \frac{\exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})}{\sum_{k \in \mathcal{J}} \exp(x_{kt}\beta - \alpha p_{kt} + \xi_{kt})} \right) \end{aligned}$$

Where the second line follows from our derivations above and using N people in the sample. Now, when a firm changes the price of a good it will not only affect the demand for their own good but also impact their own portfolio. If I'm Apple and I make the iPhone cheaper, less people will buy iPods (do those even exist anymore??):

$$\begin{aligned} \frac{\partial \pi_{jt}}{\partial p_{j't}} &= -\alpha \pi_{j't}(1 - \pi_{j't}) && \text{if } j = j' \\ &= \alpha \pi_{jt} \pi_{j't} && \text{otherwise for } j \neq j' \end{aligned}$$

Whenever firm changes price of good j'. We can now use these to derive the change in profits:

$$\frac{\partial \Pi_{ft}}{\partial p_{j'}} = \pi_{j't} + \alpha \pi_{j't} \left[\sum_{j \in J} \sum_{j \neq j'} (p_{jt} - mc_{jt}) \pi_{jt} - (p_{jt} - mc_{jt})(1 - \pi_{jt}) \right]$$

So now define $\mathcal{F}_t$ as the set of products in a firms portfolio and the elements of the FONC matrix can be written now for all goods j:

$$\frac{\partial \Pi_{ft}}{\partial p_{j'}} = \pi_{j't} + \alpha \pi_{j't} \left[\sum_{j \in J} \sum_{j \neq j'} 1\{j \in \mathcal{F}_t\}(p_{jt} - mc_{jt})\pi_{jt} - (p_{jt} - mc_{jt})(1 - \pi_{jt}) \right] = 0$$

$$\frac{\partial \Pi_{ft}}{\partial p_{j'}} = \pi_{j't} + (\Delta \cdot \mathcal{J}_p)'_{.j'}(p - mc) = 0$$

Where Δ is a matrix of 0 and 1s signallying whether or not the firm has product j in their portfolio in the specific market. Defining $\mathcal{J}_p = -\alpha \mathcal{J}_{\xi \to s}$. We can now collect this term for all firms:

$$\pi_{.t} + (\Delta \cdot \mathcal{J}_p)'(p - mc) = 0$$

Which rearranges to:

$$p = mc - (\Delta \cdot \mathcal{J}_p)^{-1} \pi_{.t}$$

. Which is the fixed point equation from before.

> **8.** Open the data set. Consider market $t = 17$. Ignore cost data, setting $MC = 0$, simulate values for ξ such that $\xi \sim \mathcal{N}(0, 1)$. Use $\theta = [-3, 1, 1, 2, -1, 1]$ to obtain the share equation for each good and solve the fixed-point equation for the optimal prices. Interpret these prices as markups, and describe any patterns at the firm level. Is this related to IIA?

From here on out, please refer to our jupyter notebooks.

To solve for the fixed point let's rearrange

$$0 = mc - (\Delta \cdot \mathcal{J}_p)^{-1} \pi_{.t} - p$$

Setting $mc = 0$

$$0 = -(\Delta \cdot \mathcal{J}_p)^{-1} \pi_{.t} - p$$

And remembering that the Jacobian is a j x j matrix where each element on the diagonal is

$$-\alpha \pi_{j't}(1 - \pi_{j't}), \text{ if } j = j'$$

And the off diagonals are:

$$\alpha \pi_{jt} \pi_{j't}, \text{ otherwise for } j \neq j'$$

For the results please see the attached Jupyter notebook. However we see that all companies have prices > 0 meaning that there is a markup over the $mc = 0$ assumption. Furthermore, the brand with the higher Xs get higher markups which make sense in a mc= 0 world as they have the best product. We also see that IIA does hold in this model if you look back to the answer in question 2.

> Now we allow for preference heterogeneity, so the preference parameters are β_i and α_i. Assume the following distribution
>
> $$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \bar{\alpha} \\ \bar{\beta} \end{pmatrix}, \text{diag}\left(\sigma_\alpha^2, \sigma_\beta^2\right) \right) \quad \boldsymbol{\theta} = [\alpha, \boldsymbol{\beta}, \boldsymbol{\sigma}]$$
>
> For identification purposes, only σ_α and $\sigma_{\beta cc}$ are non-zero. (cc denotes the characteristic labeled engine size).

9. Write a share prediction function sHat (Do not be afraid to use multiple functions). This function should take $\boldsymbol{\delta}, X, \boldsymbol{\sigma}, \zeta, I$ as inputs and return a vector of predicted shares $\hat{s}$. Verify this function for these test cases

- $\boldsymbol{\sigma} = 0, \boldsymbol{\delta} = 0, J = 3$
- $\boldsymbol{\sigma} = 0, J = 3, \delta_1 = 40, \boldsymbol{\delta}_{-1} = 20$. (Shares should not be one)
- $\boldsymbol{\sigma} = .1, \boldsymbol{\delta} = 0, J = 3, I = 20, X = 0$. How much did this change your shares? Try for other values of X

Under preference heterogeneity, we cannot aggregate consumers as we did before. Now, we have to integrate over individuals in order to get the shares. Since we are given that α_i and β_i are jointly normal, we can rewrite as follows. Define $\lambda_i = \begin{bmatrix}\alpha_i \\ \beta_i\end{bmatrix}$:

$$\lambda_i = \bar{\lambda} + F'\zeta_i$$

where F is the cholesky decomposition of the variance matrix $\Sigma_\lambda = \text{diag}(\sigma_\alpha^2, \sigma_\beta^2)$ and ζ_i is a vector where each element is distributed $\mathcal{N}(0,1)$. Then we can rewrite our mean utility (after taking the expectation over shocks $_{ijt}$ to be:

$$\begin{aligned} \bar{u}_{ijt} &= X_{jt}(\bar{\lambda} + \zeta_i \sigma) + \xi_{jt} \\ \text{Define } \delta_{jt} &= X_{jt}\bar{\lambda} + \xi_{jt} \\ \implies \bar{u}_{ijt} &= \delta_{jt} + X_{jt}\zeta_i\sigma \end{aligned}$$

Then we can write the share equation as:

$$\begin{aligned} s_{jt} &= \int \mathbb{P}(i \to j) dF(\zeta_i) \\ &= \int \frac{\exp(\bar{u}_{ijt})}{1 + \sum_{k \in \mathcal{J}} \exp(\bar{u}_{ikt})} dF(\zeta_i) \end{aligned}$$

We compute shares by doing the following:

1. Writing a function to calculate individual choice probabilities $\frac{\exp(\bar{u}_{ijt})}{1+\sum_{k \in \mathcal{J}} \exp(\bar{u}_{ikt})}$
2. Writing a function to calculate $\bar{u}_{ijt}$ for each individual
3. Wrapping the two functions above in the function $sHat$ to loop through I individuals and compute the individual choice probabilities. Then, this function integrates by taking the mean across individuals (Monte Carlo integration).

We go through several test cases in addition to the ones described above. All the computation here is done within Market 17

- **Test Case 4**: Same as Test Case 3 but X is not 0 and is set equal to the market 17 values (for the first $J = 3$ observations)
- **Test Case 5** Like Test Case 2 but weith $\delta_1 = 4, \delta_{-1} = 2$. This provides a less extreme share distribution than Test Case 2 and will allow us to test the contraction mapping in the next section.

The results are depicted below.

Figure 1: Share Vectors Corresponding to Test Cases 1 through 5

```
Test 1:   [0.25 0.25 0.25]
Test 2:   [9.99999996e-01 2.06115361e-09 2.06115361e-09]
Test 3:   [0.25 0.25 0.25]
Test 4:   [0.2495407  0.22995882 0.2362604 ]
Test 5:   [0.77580349 0.10499359 0.10499359]
```

The results generally make sense. For example, in test cases 2 and 5 where the mean utility δ is much higher for $J = 1$ than all other products, we see that product 1 has a much higher market share than all other products. In the other 3 test cases, the shares are more equal between the three products and the outside option. In Test Case 1 there is no heterogeneity in consumer preferences since $\sigma = 0$) and since all the δ elements are equal, shares are exactly equal. In Test Case 3, there is additional heterogeneity but since $X = 0$, it does not matter. Finally, in Test Case 4, when we allow non-zero X there is a small amount of heterogeneity in the shares for each good.

> **10.** Write a share inversion function (Do not be afraid to use multiple functions). This function should take $s, X, I, \boldsymbol{\sigma}, \zeta, J$ as inputs and return a δ such that $\hat{s}(\delta, X, \boldsymbol{\sigma}, \zeta, I) = s$ Verify this function using the shares that were output from your test cases previously.

We use the contraction mapping method on the equation

$$\boldsymbol{\delta}_t^{h+1} = \boldsymbol{\delta}_t^h + \log \boldsymbol{s}_t - \log\left(\hat{s}\left(\boldsymbol{\delta}_t^h\right)\right),$$

where $\hat{s}(\delta_t^h)$ is computed using the share calculating function from Question 9. We verify the results using the shares from Test Case 5 and find that the solver converges in 1949 iterations (at tolerance level $1e^{-14}$) to the vector $\delta = 0$.

> **11.** Write the objective function for the estimation problem in terms of an implicit function $\xi(X, \theta, s)$. Write the function so that it can take any arbitrary weighting matrix W. What is the objective value for $\boldsymbol{\theta} = 0$, using the weighting matrix $W = (ZZ')^{-1}$?

We write the objective function to:

1. Split the vector θ into mean and variance components
2. Estimate δ^*, the value of δ that best matches predicted and actual shares, using the inversion function from Question 10
3. Calculate the structural residual $\xi(X, \theta, s)$ using $\xi = \delta^* - X\bar{\lambda}$ (recall $\bar{\lambda} = \begin{bmatrix} \bar{\alpha} \\ \bar{\beta} \end{bmatrix}$
4. Return the objective function: $M(\theta) = (Z'\xi)'W(Z'\xi)$

Note that since we need to calculate ξ in other parts (e.g. for gradients), I create a separate **get_xi** function that just does steps 2 and 3. Also note that in order to make sure that the objective function return was not overly sensitive to the starting guess for δ, I had to set the contraction mapping tolerance to 1e-20.

Using a vector of zeroes for θ and the weighting matrix given in the problem, we compute the objective function for market 1 and get a value of 59.1979. Note that we later write an optimizer that would ideally loop through all markets to get ξ_j for each one and then feeds a concatenated ξ vector to the objective function

> **12.** Write a gradient function. (Do not be afraid of making this multiple functions). Verify your gradient using Forward-Stepwise Automatic differention. If your programming language does not support AD, use finite differences, but expect error. For Julia, use ForwardDiff; for Python use Jax.

The objective function is:

$$\min_{\bar{\lambda},\sigma} \left(Z'\xi(\theta)\right)' W \left(Z'\xi(\theta)\right)$$

We know that $\xi(\theta) = \delta(\sigma) - X\bar{\lambda}$, so we can rewrite the objective as:

$$\min_{\bar{\lambda},\sigma} \left(Z'(\delta(\sigma) - X\bar{\lambda})\right)' W \left(Z'(\delta(\sigma) - X\bar{\lambda})\right)$$

Taking the FOC with respect to $\bar{\lambda}$, we see that our optimal values for λ can be written as a function of σ solving:

$$\bar{\lambda}^* = \left(X'ZWZ'X\right)^{-1} X'ZWZ'\delta(\sigma)$$

So the objective function is now:

$$\min_{\sigma} \left(Z'\xi(\sigma)\right)' W \left(Z'\xi(\sigma)\right)$$

Following the notes from the TA session, we define the following functions to differentiate the objective:

$$\begin{aligned} f(x) &= x'Wx \\ g(y) &= Z'y \\ M(\sigma) &= f(g(\xi(\sigma))) \end{aligned}$$

The derivatives can be written out as:

$$\begin{aligned} \nabla f &= 2J_g'Wg(\sigma) \\ J_g &= Z'J_y \\ \implies \nabla f &= 2(Z'J_y)'Wg(\sigma) \end{aligned}$$

What we need to derive is J_y which is equal to $\frac{\partial\xi}{\partial\sigma}$ in our current setup. To do this, let us look at the share equation and solve implicitly:

$$\begin{aligned} 0 &= s_k - \underbrace{\left[\frac{1}{I}\sum_{i=1}^{I} \frac{\exp\left(x_{jt}\left(\bar{\lambda} - \sigma\zeta_i\right) + \xi_{jt}\right)}{1 + \sum_k \exp\left(X_{kt}\left(\bar{\lambda} - \sigma\zeta_i\right) + \xi_{kt}\right)}\right]}_{\hat{s}} \\ 0 &= \underbrace{\frac{\partial\hat{s}}{\partial\bar{\lambda}}\frac{\partial\bar{\lambda}}{\partial\sigma}}_{0 \text{ by env. thm.}} + \frac{\partial\hat{s}}{\partial\sigma} + \frac{\partial\hat{s}}{\partial\xi}\frac{\partial\xi}{\partial\sigma} \\ \implies \frac{\partial\xi}{\partial\sigma} &= \left(\frac{\partial\hat{s}}{\partial\xi}\right)^{-1}\frac{\partial\hat{s}}{\partial\sigma} \end{aligned}$$

Now let us solve for $\frac{\partial \hat{s}_j}{\partial \xi_j} \equiv J_\xi$. Define $m_{ij} = X_{jt}(\bar{\lambda} - \sigma\zeta_i) + \xi_{jt}$. From the share equation, using the quotient rule:

$$\begin{aligned}
&\frac{(1+\sum_k \exp(m_{ik}))\exp(m_{ij}) - \exp(m_{ij})\exp(m_{ij})}{(1+\sum_k \exp(m_{ik}))^2} \\
&= \frac{\exp(m_{ij})\left[1+\sum_k \exp(m_{ik}) - \exp(m_{ij})\right]}{\left(1+\sum_k \exp(m_{ik})\right)^2} \\
&= \hat{s_j}(1-\hat{s}_j)
\end{aligned}$$

Following a similar calculation, we get that $\frac{\partial \hat{s}_j}{\partial \xi_n} = -\hat{s}_j \hat{s}_n$ for $n \neq j$. So we have:

$$\frac{\partial \hat{s}_j}{\partial \xi_n} = \begin{cases} \hat{s}_j\left(1-\hat{s}_j\right) & \text{for } j = n \\ -\hat{s}_j \hat{s}_n & \text{for } j \neq n \end{cases}$$

Next let us solve for $\frac{\partial \hat{s}_j}{\partial \sigma} \equiv J_\sigma$. Let us rewrite individual choice probabilities using the function

$$h_j(m_{i1}, \ldots, m_{iJ}) = \frac{\exp(m_{ij})}{1+\sum_k \exp(m_{ik})}$$

Define the index $l \in \{1, \ldots, L\}$ to represent of coefficients in the model. Using the same quotient rule method as above, the derivatives of the function h_j are as follows:

$$\begin{aligned}
\frac{\partial h_j}{\partial m_{ij}} &= \pi_{ij}(1-\pi_{ij}) \\
\frac{\partial h_j}{\partial m_{in}} &= -\pi_{ij}\pi_{in}
\end{aligned}$$

where π_{ij} is the choice probability $\mathbb{P}(i \to j)$. Note also that $\frac{\partial m_{ij}}{\partial \sigma} = X_{jt}\zeta_i$

$$\frac{\partial \hat{s_j}}{\partial \sigma_l} = \sum_{n=1}^{J} \frac{\partial h_j}{\partial x_n} X_{nt,l}\zeta_{i,l}$$

Writing this in vector notation:

$$\begin{aligned}
\left(\frac{\partial \hat{s}_j}{\partial \sigma}\right)_i &\equiv J_\sigma^i = \nabla h_j^i X_t \operatorname{diag}(\zeta_i) \\
\implies \frac{\partial \hat{s}_j}{\partial \sigma} &= \frac{1}{I}\sum_{i=1}^{I} \nabla h_j^i X_t \operatorname{diag}(\zeta_i)
\end{aligned}$$

So then we have

$$\begin{aligned}
J_y = \frac{\partial \xi}{\partial \sigma} &= (J_\xi)^{-1} \frac{1}{I}\sum_{i=1}^{I} J_\sigma^i \\
\implies \nabla f &= 2\left(Z'\left(J_\xi\right)^{-1} \frac{1}{I}\sum_{i=1}^{I} J_\sigma^i\right)' W Z' \xi(\sigma)
\end{aligned}$$

See the attached Jupyter notebook for the code for the gradient. We do it in 2 pieces: calculating the J_y matrix and then calculating the full marix. We verify against Jax.

13. Estimate your results using Two-stage GMM, beginning with the TSLS weighting matrix. Report point estimates as well as the standard errors.

Our two-stage GMM consists of the following steps:

1. Take a guess of the weighting matrix (we use the identity with dimension equal to the number of instruments)
2. Pass the weighting matrix to the optimizer function (see details below, this nests the objective function)
3. Use the resulting output of θ to calculate the optimal weighting matrix:

$$\frac{1}{J} Z'\xi(Z'\xi)'$$

   where L is the number of instruments, Z' is $L \times J$ and ξ is $J \times 1$. This is like the "square" of the GMM sample moment condition averaged across observations J

4. Rerun the optimization with the optimal weighting matrix to get final parameter estimates

Some notes on the optimizer function (**gmm_optimize**). This function is a wrapper for our objective function that does the following:

1. Calculates ξ, δ using the **get_xi** function that uses contraction mapping
2. Defines a minimizer and a minimizer gradient function that only take σ as an argument
3. Optimizes using the BFGS method in scipy
4. Calculates the final value of δ using the contraction mapping function but this time feeding in σ^* from the optimization
5. Calculating $\bar{\lambda}^*$ as described in the problems above

We note that we were only able to get this to run on a singular market (Market 1). We include alternative versions of the optimizer and gmm functions above (labeled **gmm_optimize_alt** and **gmm_alt**) at the end of our Jupyter notebook. In the alternative setup, we cycle through all markets to calculate shares and ξ_j for each one and then pass the whole vector of ξ (concatenated across markets) to to the optimizer. We were unable to get this to converge in a reasonable amount of time, however (the code has been running for 6 hours). Below we report GMM parameter estimates for just Market 1.

$$\bar{\lambda} = \begin{bmatrix} 0.3859 & -0.3225 & 4.3034 & -1.7641 & -4.6807 & 1.1430 \end{bmatrix}$$
$$\sigma = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We doubt that our GMM estimates are spot on for the reasons described above. Our estimates imply that the random coefficients specification is entirely unnecessary and there is no heterogeneity in consumer preferences. Our estimates are off probably due to the fact we ran on a market with less observations than free parameters. Due to lack of time we did not provide estimates for the full sample we just included code on how to do so. Comparing to the σ values found from pyBLP, however, it does seem that there is not too much variation in preferences.

14. Compare your results to pyBLP (This does not need to be done in python, you can call pyBLP from R, matlab, and Julia.)

See Jupyter Notebook. We find a negative coefficient (-3 to -4) on prices and a positive coefficient on engine prices which is exactly what we'd expect so we are feeling good. (this number depends on the zeta shocks). Additionally we find positive coefficients on EngineSize and SportsBike as well as low sigmas.

```
=============================================================================

Nonlinear Coefficient Estimates (Robust SEs in Parentheses):
===============================================================================================
  Sigma:        prices          EngineSize       |   Pi:        alpha_shock      engine_shock
----------  ---------------  ---------------     |  ----------  ---------------  ---------------
  prices     +1.726416E-02                        |   prices     +9.491797E-01    +0.000000E+00
            (+3.204131E+00)                      |             (+6.342611E+00)
                                                 |
EngineSize   +0.000000E+00    +4.316202E-01       |  EngineSize   +0.000000E+00    +9.631692E-01
                             (+2.045751E+00)     |                              (+6.506126E+00)
===============================================================================================

Beta Estimates (Robust SEs in Parentheses):
==================================================
    prices         EngineSize       SportsBike
---------------  ---------------  ---------------
 -3.990281E+00    +1.250046E+00    +2.395143E+00
(+1.234147E+01)  (+3.209621E+00)  (+4.327380E+00)
==================================================
```

15. For market $t = 17$, print the elasticity matrix. Is IIA still present? How close are we to IIA? Was this worth the effort of adding hetereogeneity? Feel free to use your supply-side results from the homogenous section.

```python
elasticities = results.compute_elasticities().round(decimals = 2)
print(np.round(elasticities[17*7+1:17*7+8,0:6]))
```

```
Computing elasticities with respect to prices ...
Finished after 00:00:00.

[[-6.  1.  0.  0.  0.  4.]
 [ 0. -2.  0.  0.  0.  4.]
 [ 0.  1. -4.  0.  0.  4.]
 [ 0.  1.  0. -9.  0.  4.]
 [ 0.  1.  0.  0. -8.  4.]
```

We do appear to still have some degree of IIA in this model as it is still a logit-based model however, the heterogeneity buys us a little wiggle room. In this regard it was worth adding the heterogeneity - as it fits the real world better. The small sigmas indicated above show we do get some benefit from heterogeneity but probably not worth all that effort honestly.